

Supplementary Table 1: Baseline characteristics of the study participants

Characteristics	Basic prediction model (N=608)		Advanced prediction model (N=540)	
	Training set, n (%) N=490	Validation set, n (%) N=118	Training set, n (%) N=432	Validation set, n (%) N=108
Mean age (±SD)	52.9 (17.5)	53.6 (15.7)	54.1 (16.8)	55.5 (18.1)
Sex				
Male	306 (62.4)	80 (67.8)	288 (66.7)	59 (54.6)
Female	184 (37.6)	38 (32.2)	144 (33.3)	49 (45.4)
Education (years of schooling)				
No formal education	54 (11.0)	12 (10.2)	50 (11.6)	16 (14.8)
Primary (1-5)	58 (11.8)	18 (15.2)	60 (13.9)	16 (14.8)
Secondary and Higher secondary (6-12)	126 (25.7)	28 (23.7)	117 (27.1)	33 (30.6)
Graduate and above	252 (51.5)	60 (50.9)	205 (47.4)	43 (39.8)
Occupation				
Employed	234 (47.8)	58 (49.1)	209 (48.4)	46 (42.6)

Unemployed	256 (52.2)	60 (50.9)	223 (51.6)	62 (57.4)
Comorbidities				
Diabetes Mellitus				
Present	148 (30.2)	45 (38.1)	151 (34.1)	35 (32.4)
Absent	342 (69.8)	73 (61.9)	281 (65.9)	73 (67.6)
Hypertension				
Present	117 (23.9)	36 (30.5)	120 (27.8)	30 (27.8)
Absent	373 (74.1)	82 (69.5)	312 (72.2)	78 (72.2)
Chronic Kidney disease				
Present	8 (1.6)	4 (3.4)	9 (2.1)	3 (2.8)
Absent	482 (98.4)	114 (96.6)	423 (97.9)	105 (97.2)
Chronic Obstructive Pulmonary disease				
Present	4 (0.8)	0 (0)	2 (0.5)	1 (0.9)
Absent	486 (99.2)	118 (100)	430 (99.5)	107 (99.1)
Bronchial asthma				
Present	3 (0.6)	2 (1.7)	2 (0.5)	2 (1.8)

Absent	487 (99.4)	116 (98.3)	430 (99.5)	106 (98.2)
Tuberculosis				
Present	4 (0.8)	1 (0.8)	2 (0.5)	2 (1.8)
Absent	486 (99.2)	117 (99.2)	430 (99.5)	106 (98.2)
Coronary Artery disease				
Present	37 (7.6)	4 (3.4)	29 (6.7)	11 (10.2)
Absent	453 (92.4)	114 (96.6)	403 (93.3)	97 (89.8)
Hypothyroidism				
Present	26 (5.3)	4 (3.4)	24 (5.4)	4 (3.7)
Absent	464 (94.7)	114 (96.6)	408 (94.6)	104 (96.3)
Behavioral habits				
Tobacco use				
Present	33 (6.7)	11 (9.3)	34 (7.9)	6 (5.6)
Absent	457 (93.3)	107 (90.7)	398 (92.1)	102 (94.4)
Alcohol use				
Present	36 (7.4)	14 (11.9)	39 (9.0)	7 (6.5)

Absent	454 (92.6)	104 (88.1)	393 (91.0)	101 (93.5)
Symptom profile				
Fever				
Present	338 (69.0)	86 (72.9)	318 (73.6)	75 (69.4)
Absent	152 (31.0)	32 (27.1)	114 (26.4)	33 (30.6)
Cough				
Present	351 (71.6)	92 (78.0)	331 (76.6)	76 (70.4)
Absent	139 (28.4)	26 (22.0)	101 (23.4)	32 (29.6)
Dyspnea				
Present	187 (38.2)	51 (43.2)	187 (43.3)	46 (42.6)
Absent	303 (61.8)	67 (56.8)	245 (56.7)	62 (57.4)
Olfactory-gustatory dysfunction (loss of taste/loss of smell)				
Present	97 (19.8)	23 (19.5)	64 (14.8)	20 (18.5)
Absent	393 (80.2)	95 (80.5)	368 (85.2)	88 (81.5)
Gastrointestinal symptoms (vomiting/diarrhea)				
Present	99 (20.2)	18 (15.3)	81 (18.8)	22 (20.4)

Absent	391 (79.8)	100 (84.7)	351 (81.2)	86 (79.6)		
Biochemical parameters						
C-reactive protein (mg/L)						
<100			356 (82.4)	94 (87.0)		
≥100			76 (17.6)	14 (13.0)		
D-dimer (ng/mL)						
<1000			333 (77.1)	83 (76.8)		
≥1000			99 (22.9)	25 (23.2)		
Ferritin (mcg/L)						
<500			270 (62.5)	74 (68.5)		
≥500			162 (37.5)	34 (31.5)		
Lactic acid dehydrogenase (units/L)						
<400			201 (46.5)	52 (48.1)		
≥400			231 (53.5)	56 (51.9)		
Radiological parameter						
CT severity score (% lung involvement)						

No involvement		29 (6.7)	8 (7.4)
<25%		99 (22.9)	29 (26.9)
25-50%		174 (40.3)	44 (40.7)
51-75%		100 (23.2)	18 (16.7)
>75%		30 (6.9)	9 (8.3)

Supplementary Table 2: Multivariable logistic regression for the predictors of COVID-19 severity based on training set – basic prediction model (N=490)

Characteristics	Adjusted OR	S.E.	95% CI	p-value
Age of the patient	1.03	0.008	1.01-1.05	0.001*
Sex of the patient				
Male	1.96	0.47	1.22-3.13	0.005*
Female	Ref	-	-	-
Education of the patient				
No formal education	2.84	1.14	1.29-6.24	0.01*
Primary	0.95	0.37	0.44-2.02	0.89
Secondary and Higher secondary	0.85	0.27	0.46-1.58	0.61
Graduate and above	Ref	-	-	-
Tobacco use				

Present	3.56	1.65	1.43-8.86	0.006*
Absent	Ref	-	-	-
Chronic Kidney Disease				
Present	17.70	20.26	1.88-166.88	0.01*
Absent	Ref	-	-	-
Cough				
Present	2.25	0.62	1.31-3.86	0.003*
Absent	Ref	-	-	-
Dyspnea				
Present	3.31	0.74	2.13-5.12	<0.001*
Absent	Ref	-	-	-
Olfactory-gustatory dysfunction (loss of taste/loss of smell)				
Present	0.47	0.15	0.25-0.87	0.02*

Absent	Ref	-	-	-
Gastrointestinal symptoms (vomiting/diarrhea)				
Present	1.90	0.53	1.10-3.29	0.02*
Absent	Ref	-	-	-

OR-Odds Ratio; S.E-Standard Error; CI-Confidence Interval; Ref-Reference category

*p-value <0.05 i.e., statistically significant

Supplementary Table 3: Multivariable logistic regression for the predictors of COVID-19 severity based on training set – advanced prediction model (N=432)

Characteristics	Adjusted OR	S.E.	95% CI	p-value
Age of the patient	1.02	0.009	1.01-1.04	0.01*
Sex of the patient				
Male	1.76	0.49	1.01-3.05	0.04*
Female	Ref	-	-	-
Education of the patient				
No formal education	2.54	1.14	1.05-6.11	0.04*
Primary	0.81	0.33	0.36-1.82	0.62
Secondary and Higher secondary	0.77	0.27	0.38-1.54	0.45
Graduate and above	Ref	-	-	-
Tobacco use				

Present	3.16	1.59	1.18-8.46	0.02*
Absent	Ref	-	-	-
Chronic Kidney Disease				
Present	21.44	27.10	1.80-255.39	0.01*
Absent	Ref	-	-	-
Cough				
Present	2.26	0.74	1.19-4.30	0.01*
Absent	Ref	-	-	-
Dyspnea				
Present	2.13	0.53	1.30-3.47	0.003*
Absent	Ref	-	-	-
Gastrointestinal symptoms (vomiting/diarrhea)				
Present	2.39	0.74	1.31-4.38	0.005*

Absent	Ref	-	-	-
C-reactive protein (mg/L)				
<100	Ref	-	-	-
≥100	2.35	0.77	1.24-4.47	0.009*
D-dimer (ng/mL)				
<1000	Ref	-	-	-
≥1000	1.97	0.57	1.12-3.46	0.01*
Ferritin (mcg/L)				
<500	Ref	-	-	-
≥500	1.69	0.44	1.01-2.82	0.04*
Lactic acid dehydrogenase (units/L)				
<400	Ref	-	-	-
≥400	1.83	0.48	1.10-3.04	0.02*

CT severity score (% lung involved)				
<50%	Ref	-	-	-
≥50%	2.69	0.72	1.59-4.54	<0.001*

OR-Odds Ratio; S.E-Standard Error; CI-Confidence Interval; Ref-Reference category

*p-value <0.05 i.e., statistically significant